UNITED NATIONS

ECONOMIC
AND
SOCIAL COUNCIL

LIMITED
CLADES/WG.1/L.13
For participants only
24 August 1971
ORIGINAL: ENGLISH

STRUCTURE AND FLEXIBILITY IN CLASSIFICATION SCHEMES FOR DOCUMENTATION: A PROPOSAL FOR LATIN AMERICA

Prepared for the Comité Consultivo of the Centro Latinoamericano de Documentación Económica y Social (CLADES), and the Meeting on Modern Documentation Techniques, by Win Crowther, of the Economic Commission for Latin America (ECLA).

71-9-2515

INDEX

1. Introduction

1. The purpose of this article is to describe the necessity of balancing two goals, those of structure and flexibility, in the design of a classification scheme for social and economic documentation in Latin America. Several alternatives will be described, and the implications for automatic data retrieval and relevance to user needs will be indicated. It is hoped that the reader will be impressed that the problem is a complex one, but that means must be found by which the users of the documentation system can intervene in the final choice of a classification scheme.

2. It is important to stress at the outset that the field of documentation is in a great deal of flux. It is a profession that has passed through a succession of sets of orientating concepts, without a clear resolution of the major issues. There is a great deal of experimentation and innovation, especially oriented to deal with the major problems that have beset existing documentation schemes, such as the incorporation of new classifications, the very high cost of retrieval, the problem of interchanging documentation among the various documentation centers, and the lack of relevance to the information needs of the users. I have not read anything by any widely-recognized or competent expert in the field indicating that any one system is "proven" to be superior for international application. On the contrary, this profession is distinguished because there is a new strong desire within it to find "solutions" which are appropriate to the particular users and their needs in each situation. 1/ There is also a great desire to find ways to increase

---

1/ In this respect, Eric de Grolier a leader in the field for over 30 years, reported in 1965 that "This idea came to me last night - that it was very important to study the questions posed by the users and to deduct from those questions the useful classifications". "Current Trends in Theory and Practice of Classification", in Classification Research, ed. by Pauline Atherton (Copenhagen: Munksjaard, 1965), page 12. Lancaster articulates the user emphasis in these terms: "Efficient subject indexing is not necessarily achieved by labeling a document on the basis of its intrinsic subject matter. Rather, it is achieved by labeling a document according to the types of users who may be expected to derive most benefit from it, and the types of requests for which the document is likely to be regarded as responsive... The same document may quite correctly be indexed in six completely different ways in six separate organizations." F. Wilfrid Lancaser, Information Retrieval Systems (New York, 1968), p.2.

/the compatability

the compatability among documentation systems, facilitating the interchange of work among documentation centers, but this is a secondary goal. It is hoped that new computer technology will facilitate the automatic reclassification of materials, and all who work in the field take this into consideration during the design of classification schemes.

3. In this paper we will consider the efficiency of a documentation system to be the amount of correspondence between (a) the material in a given output as a result of a request and (b) the material that the requester would have selected had he scanned completely every text documented in the system. The goal is (a) to include a very high proportion of the "relevant" materials (the recall ratio) and (b) to exclude a very high proportion of the materials which are not "relevant" to the user (the precision ratio). A trade-off is almost always necessary between recall and precision, and economic cost must be taken into consideration.

4. We will suggest that the most appropriate system for the documentation of Latin American social and economic information would probably incorporate a superimposition of rigid and flexible systems. Rigid systems are usually used where precision is stressed and flexible systems tend to be used where recall is stressed. No pretension is made that finding a suitable compromise is an easy task and our proposal here is only tentative. As we investigated more material on documentation, we have reformulated our ideas. We will continue this education and reformulation, and hope that for this we will receive guidance from all persons interested in this subject.

2. Why flexibility is desirable and undesirable

5. The more that one knows about a field, and the less rigid he is in his commitment to certain concepts, the more likely he will prefer maximum flexibility in the classification of information in his field. J.R. Smith has stated this in these terms:

> "...it was suggested that hierarchical classification dated rapidly and soon became inadequate, and that in general universal classifications appeared to be of most use to scientists in the fields other than their own." 2/

2/ J.R. Smith, "The feasibility of a world system: UNISIST by ICSU our of UNESCO", Aslib Proceedings (August, 1970), page 397.

/6. Superficially the

6. Superficially the idea of flexibility is very attractive. One wants the right to add and delete or to modify the meaning associated with classification titles (index terms, descriptors, keywords, etc.) as a field changes and some concepts and variables take on new importance. Also, one wants the freedom to draw from a wide variety of classification titles when describing a document. It is easily seen that the essense of what an article is about, and the relative importance of the topics treated in the article, can be lost (or butchered) if the documentalist is told to draw from a very small or highly structured set of classification titles.

7. In recognition of these advantages of flexibility, several international agencies, led by the International Labour Organization (ILO), and later the Organization for Economic Cooperation and Development (OECD), sponsored the development (1) of a very extensive list of social and economic descriptors, which could easily be modified (although the documents described previous to such modifications would, of course, not be described by the new terms incoporated into the list), and (2) of a synopsis form, which would allow a great deal of freedom in describing the document. The OECD had many years of experience in this field, especially in documentation in science and technology, and was partisan of great flexibility. This is evident in comments by OECD experts in this field and in the proposed list of social and economic descriptors, the Aligned Descriptor List (ADL).

8. In reviewing the work of the OECD, the Information Policy Group of that organization has very recently made an incisive self evaluation, recognizing disadvantages of the extreme position they had taken on the flexibility-structure issue:

> Mr. P.J. Judge, Head of the Section for Scientific and Technical Information, Directorate for Scientific Affairs, OECD, states that the Information Policy Group has been: "satisfied that a policy for information should include the provision of 'information for policy' - the scientific and technical information needed by government officials concerned with a wide variety of problems, as part of the input to the formulation of policy . . . The IPG would also now place much greater emphasis on the development of inter-disciplinary working; on the development of information

/services in

services in the social sciences, and the integration of these into multidisciplinary systems including information and data from the natural sciences; and on the education and training aspects, looking to the provision of new kinds of information specialists to cope, in a 'user-oriented' way, with the increasing complexity of the information problems which are arising and the increasing sophistication of the techniques becoming available to resolve them." 3/

9. Specifically with regard to the Aligned Descriptor List, it was recognized by the experts working on that list, at the end of 1969, that the first edition of the ADL was "inadequate", and that a reform such that "gaps could be filled, obvious incompatibilities settled and definitions given so as to clarify the use of an ambiguous descriptor, would be of little use. A more radical transformation would have to be considered".

10. A procedure for modifying the ADL was elaborated, aiming towards a vocabulary which would capture major elements of "most specific aspects of economic and social development", rather than just those aspects emphasized by the organizations most directly involved in the design of the new system. "To avoid any arbitrary decisions being taken, the reduction of the vocabulary should be carried out by the members of the network and correspond to a new and more strict organization of the vocabulary, the object being to obtain by the elimination of some of the already existing descriptors and the inclusion of new terms, the elements of a language comprising all the essential branches of development, and then to situate these elements in relation to each other within the framework of a real thesaurus." Evidently, the experts, who represented a number of international organizations, did not feel they were ready to consider themes that would have required even more consensus and structure than did the making of a macro-thesaurus. Thus no attempt was made even to suggest the broad outlines of an international division of documentary indexing

---

3/ P.J. Judge, "Intergovernmental plans and achievements: OECD and other international bodies", Aslib Proceedings (August, 1970), pages 371-2.

/work; whereby

work whereby organizations would not only index their own documents, but also the documents relevant to their work programme, in such a way that duplication of the analytical work would be minimized. 4/

11. Possibly because of the number of participants, the content of the macro-thesaurus being developed as of December 1970, apparently was less than that hoped for earlier, in terms of intellectual consistency and balance. An arbitrary vote-counting device was being used for deciding what terms would be included in the ADL. However, there would be a minimum amount of structuring of the thesaurus, by "grouping descriptors into appropriate clusters within relatively broad sementic field ... based on the programme classification prepared by the CCAQ (Consultative Committee on Administrative Questions [of the United Nations])". 5/

12. Outside of the OECD, as well as in it, it is becoming more and more evident that the desire for flexibility cannot lead to license for the documentalist in capturing, in a concise but literary fashion, the essense of a document. 6/ There must be a concern for the uses to which the sinopsis

---

4/ Main Conclusions, Programme of Work for 1970, of the Expert Group Meeting: "Aligned Descriptor List", Paris, 8-9 December 1969.

5/ Main Conclusions, of the Meeting of the Group of Experts, "Aligned List of Descriptors", Geneva, 7-8 December 1970. A later report which referred to this meeting suggests that "strict rules" were drawn up for introducing descriptors and also referred to the "system of classification drawn up by the United Nations Consultative Committee on Administrative Questions". Summary Record, of the Annual Meeting of the Correspondents of the Development Enquiry Service, Geneva, 9-11 December 1970.

6/ After a number of years of experience, the International Labour Organization incorporated a limited amount of hierarchy of terms in the documentation retrieval system. There will be a continuing debate on the issue of structure during the 3-5 years it will take for the Aligned Descriptors List to become fully operational. See also A.G.A. Pickford, "Some Problems of using an Unstructured Information Retrieval Language in a Coordinate Retrieval Indexing System," Aslib Proceedings 23 (March, 1971), 133-38.

will be put. And this requires a clear profile of the user and a clear conception of his information needs. The documentalist should be required to say certain things about each document, so that the user will be able to select documents on the basis of certain operating variables.

13. The Aligned Descriptors List is very extensive, but the lack of orientation has resulted in a list that is too general and too vague. Thus in the area of transport, terms like "transport facility", "transport method","flight", "navigation", and "tonnage" are useful words from a literary point of view, but they are not meaningful for the retrieval of specific bodies of information. While one cannot be extreme on the goal of having somewhat mutually exclusive terms, the ADL has so many terms that are similar and overlapping, with great variations in generality and specificity, the user will have a very difficult task in making requests that are both inclusive and exclusive to the degree he wants. Thus the user of the ADL is faced with the problem of choosing, for purposes of retrieval, between such terms as "shipping", "ships at sea", and "sea transport", in most cases finding these too general or ambiguous, yet not wanting to miss any of the material that might be printed out when reference is made to these terms. The chances are that the user will then be deluged with much more information that he really wanted, at a cost much greater than is necessary.

14. Excessive output is even more likely because the user cannot ask for one citation order rather than another, since there is no hierarchical relationship among the terms. For example, it might not be clear if one wants information on the use of "computers" in "design", or the "design" of "computers". (This will be treated in Section 9 in more detail).

15. The lack of a "policy-orientation" has meant that the list is only a compilation of terms from the separate disciplines. Thus the Aligned Descriptors List does not consider a classification of transport problems by the type of packaging or product, whereas a "policy-orientation" would certainly call for such a scheme.

16. In fact, from a purely transport point of view, this is an increasingly accepted approach to transport problems, and the exclusion of this in the ADL scheme suggests that those compiling the descriptors

/which were

which were included were not up-to-date in their fields. The same could be said for the list on public administration. However, I do not believe that this is inherent in the flexible classification, but merely a particular weakness of the ADL itself.

17. Also there is likely to be an excessive amount of matching, in the computer, of descriptors which are not directly related, but which are cited for the same document, (e.g. two terms which are related to a third term, but not to each other). Thus the mere citation of four terms: "design", locomotive", "cost", and "operation" is likely to lead to the retrieval of a great deal of unwanted information when the user only wants references on how the design of locomotives affects the cost of operations of a railroad.

18. It is true that classifications are ultimately arbitrary and should be empirical. One does, however, have to start somewhere. All experience is categorized by one means or another as soon as it is communicated (or even thought). To pretend this is not the case is to mean that a momentarily dominant set of categories in which intellectual work is divided will be taken as "empirical". In this case, the categories of the separate disciplines would be adopted, as was done when the ADL was created. There are some ways out of this dilemma. Instead of trying to categorize information by subject categories, one might use the functional relations of concepts up to a point. complementing this with the use of the flexible synopsis system. To the extent that the flexible synopsis system is incorporated, certain conditions could be established to facilitate retrieval. These possibilities will be described after some comments on the highly structured system.

3. The highly structured system

19. The highly structured system is a hierarchically arranged set of classification titles, such that each of a set of mutually exclusive categories is divided into a set of mutually exclusive subcategories, these latter are in turn divided into sets of even smaller categories, and so on. The great advantage of the highly structured system is that it provides the basis for documentation oriented to the needs of users

/at any

at any one point in time. Also, it offers a certain facility for handling documents by those who are very familiar with the classification scheme. It allows documents which have similar concerns, but which deal with these concerns using different technical terminologies, to be classified together. Through identification of the document by the lowest term in the hierarchy, the more general categories in which that document belongs are known.

20. However, the very advantages of the system imply disadvantages. A rigid structure oriented to the needs of users at any one point of time, and especially one programmed on a computer, is likely to lose relevance in the face of new concepts and variables. It requires that the immediate user of the system be familiar with the particular classification titles. It forces dissimilar documents to be crammed together under the same title because the scheme entails rigid divisions between categories of documents.

21. A somewhat less than extreme position on this is taken by the Scientific Information Exchange. 7/ The purpose here is to categorize research projects and to report to an inquiring scientist on the projects which are likely to be of greatest interest to him. A goal is to reduce to the minimum possible the need for intervention in the system or knowledge of the system by the inquiring scientist. Thus the SIE is staffed primarily by subject-oriented scientists and technicians, trained in documentation. These categorize the documents, assigning to them certain "keywords", identifying the document by one or more of the most specific keywords (in a form called "unique last term") in a hierarchical set of more general to more specific keywords under which the document is being classified (e.g. reference to "chicken" will automatically mean a case of the more generic term "poultry"). The scientist-documentalists, drawing on their close familiarity with the system and the fact that they

---

7/ D.F. Hersey, W.R. Foster, M. Snyderman, and F.J. Kreysa, "Conceptual indexing and retrieval of current research records: an analysis of problems and progress in a large scale information system". Methods of Information in Medicine (July, 1968), pages 172-87.

/have carried

have carried out the documentation in their filed of specialization, act as intermediaries between the ultimate user (the inquiring scientist) and the system. They interpret the request of the inquiring scientist and retrieve the information they believe that that scientist wants. Because of the high competence of the scientist-documentalists as interpreters of the user's requests, SIE studies indicate a very high rate of user satisfaction with the system.

22. The structure facilitates automated search and thus computer time is minimized. However the SIE finds it necessary to make a multitude of coded references to the same document to deal not only with each topic which is treated, but with all the major possible combinations of themes and citation orders. (This problem will be described in more detail later.)

4. The relations among concepts (facets) 8/

23. Most structured systems are based on the subdivision of a "universe of knowledge" according to subject headings (categories). An example is the UDC (Universal Decimal Classification). However, there are classification schemes which use titles other than subject headings, with more or less emphasis on the relations among concepts (facets). There has been a great deal of interest in this approach by documentation experts.

24. The subject heading specifies an area of social or economic activity or a type of material or task incorporated in this activity. The subject heading is generally meant to be more or less mutually exclusive with other subject headings in the same hierarchical level in the same classification scheme.

25. The use of relations among concepts for information classification assumes that all fields are really interested in systems. All intellectual exercises presume a "system", with certain boundaries and components.

8/ In this and the remainder of the discussion, references to the nature of the document itself (as opposed to the content of the document) are deleted. This is a separate problem not considered in this paper.

/"System analysis"

"System analysis" is not new. The specialty that has that name only intends to bring to the surface the presumptions involved in approaching the world or a part of the world as a "system". 2/

26. By categorizing information according to relations of concepts, it becomes evident that many fields are concerned about similar things, but in specialized terminologies which obscure these similarities. The categorization of information by disciplines has hindered the ability of a person working on a problem in one field to take advantage of the work done in other fields. The relations of concepts are those aspects of "systems" which are of common interest to a large number or all intellectual fields.

27. In the area of documentation there are a number of efforts to identify relations of concepts and use these as a basis for designing a classification scheme. It is worthwhile to review three of these. One of these, designed for the MARC project of the Library of Congress by the Classification Research Group, is much broader in interest than just social and economic documentation, while another, that of the Highway Research Board, is much more specific in interest than general social and economic documentation. The third is of the Batelle Memorial Institute.

28. The major goal of the MARC system has been described in these terms by D. Austin:

> "The aim is to bring homologous ideas together in such a way that they share a common notational base. Only in this way can an enquirer learn of work which might have some bearing on his own problems, even though it has been conducted in a different and perhaps unsuspected field of study which possesses its own terminology or jargon. It should be possible, for example, for the specialist studying the problem of, say, distributing in marketing to become aware of any relevant work on dispersion in physics or even dissemination in information science. On the understanding that distribution, dispersion and dissemination are closely related concepts, they would share enough of a common notation to make

---

2/ C. West Chruchman, The Systems Approach (New York: Dell Publishing Co., 1968).

/any one

any one of them readily retrivable to a specialist concerned in the others. So long as the notation is machine-readable, and its function is limited to the expression of this basic homologous relationship between concepts, it would then serve as a true meta-language, and should allow the specialist to scan related studies not only in other disciplines but also in other natural languages. If basic ideas can be organized in this way, it obviously should not matter whether the words which represent these concepts are in English or Swahili."

"If this system is to work, it is essential that these thesauri should be structured according to a single basic relationship between concepts which is applied consistently thoughout them. This is conceived as the generic relationship, which is already familiar in the taxonomic classification of the natural sciences, but would be applied in the proposed system on a much wider basis, to reveal not only that the horse, for example, is an equine, a mammal, a vertebrate and so on, but also that orange is a color lying in the spectral range between yellow and red, being part of a general category called appearances, and that rivets and screws belong to a category called fastening devices, and have obvious affinities with nails and adhesives." 10/

29. The designers of this system have derived the following fundamental relationships between concepts, or "operators" (using their numeration):

(0) Study region, sample population

(1) Viewpoint, perspective

(2) Active system

(3) Effect, action

(4) Key system

(5) Discipline

(6) Environment 11/

10/ D. Austin, "An information retrieval language for MARC", Aslib Proceedings (October, 1970), page 486.

11/ Unfortunately the only examples given by Austin are included in the following citation:
"These operators have two distint functions to perform in a synthetic classification. In the first place, they specify the role of the concept which follows the operator, so that, for example, an entity such as an aeroplane would be introduced by the operator (4 - Key system) if it was considered as an object in its own right and constituted the principal topic of a document, but it would be introduced by the operator (2- Active system), indicating an active system, if its special role in, say, freight or military transport were being considered. At the same time as the role of any concept is specified in this way, its citation order is also determined by a special rule of classing associated with each of the operators. As a general rule, operators are cited retroactively,which leads to an order of precedence such that a thing is set down before its parts or its properties, and also before an action performed upon it. The action must l be named before the responsible agent is cited, and so on." D.Austin, op.cit., page 487.

/The classification

30. The classification scheme of the Highway Research Board is given in Table 1. [12] There are five facets specified with a hierarchical set of subject headings under each facet. We do not have a definition of the terms "system activities", "system elements" and "system characteristics", but it can be seen that the HRB has attempted to move one step away from a strict subject classification.

31. Similar to the faceted classification indexes in many ways is the system of "roles and links", elaborated by Battelle Memorial Institute. [13] Roles are like facets in that they are syntactical devices used to display relationships between terms. They clarify, for example, if one wants information on the use of "computers" in "design", or the "design" of "computers". In addition, they may be particularly useful to social scientists and engineers who wish to have an easy-to-read description of the concepts used in studies according to the manner in which the concepts are used as variables or functions. The Battelle system also groups or "links" concepts together according to the intended association of these concepts, thus preventing the false cross association of terms belonging to separate ideas. Examples of the use of "roles and links" are shown in Table 2.

32. The faceted scheme was developed in reaction to the inconveniences of the overly flexible systems or of structured systems based on logic rather than use. [14] Also, it was realized that the problems of indexing or

---

12/ Facilitated by the Highway Research Board.

13/ J.C. Costello, Jr., Training Manual and Workbook for use in abstracting and coordinate indexing (Columbus, Ohio: Battelle Memorial Institute, 1963).

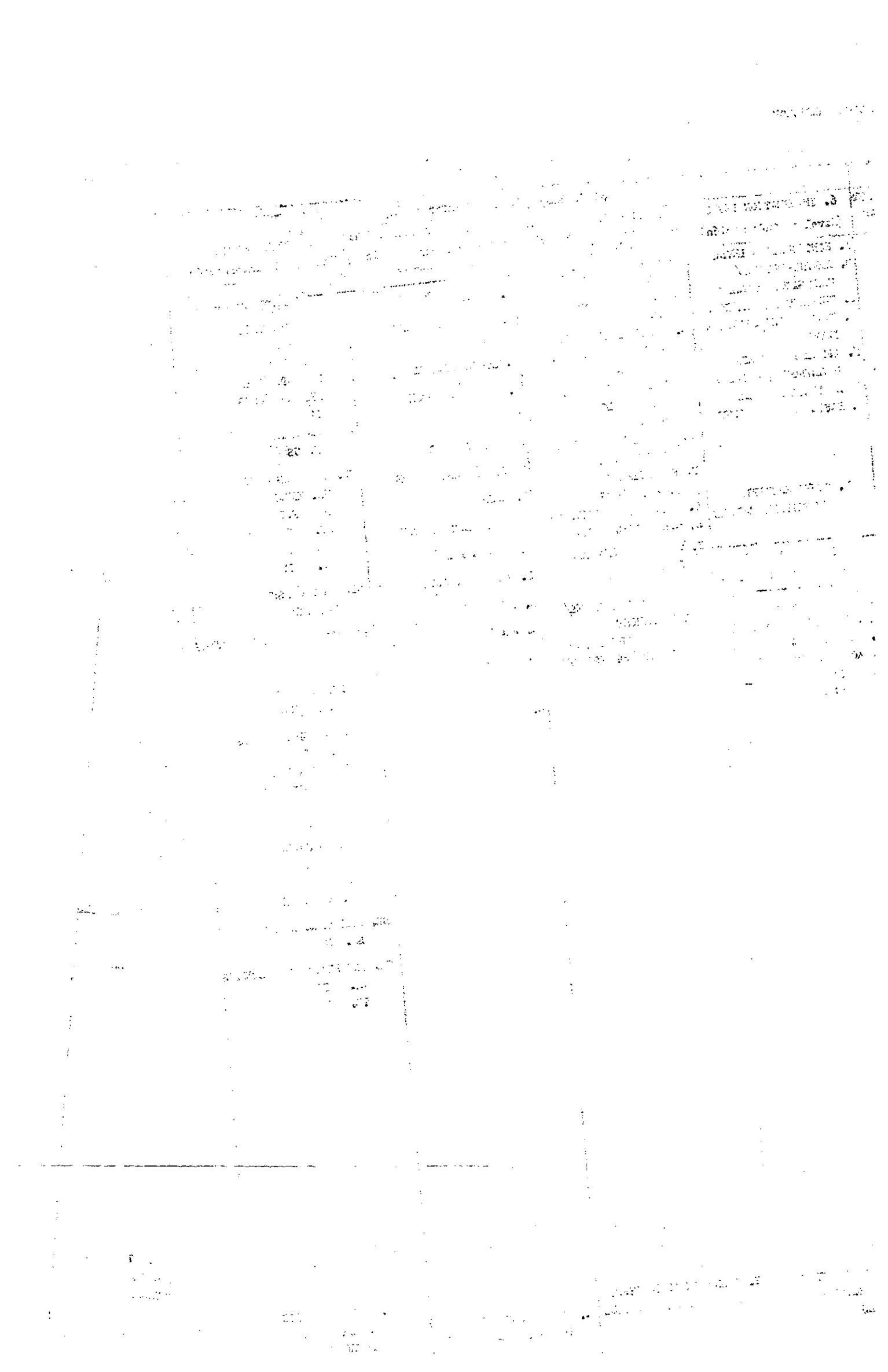
14/ Faceted schemes are, in the words of Alan Reed, "mission-oriented rather than discipline-oriented. What we mean by that is that they are designed for user groups whose interests cut across the traditional fields." Alan Reed, quoted in B.C. Vickery, Faceted Classification Schemes (New Jersey: Rutgers University Press, 1966), page 12.

/classification cannot

Table 1

INFORMATION FOR TRANSPORT RESEARCH

	RECORD CLASSES (Clases de información)			
SYSTEM CHARACTERISTICS (Características de los sistemas)	6. INFORMATION LEVEL (Nivel de información)	8. GENERATOR AGENCY TYPE (Organismos que generan información)	9. GENERATOR AGENCY LOCATION (Ubicación del organismo)	10. RECORD SOURCE (Fuentes de información)
SAFETY/ACCIDENTS NOISE/VIBRATION POLLUTION QUALITY/PERFORMANCE DURABILITY/LIFE/ RELIABILITY DEMAND/USE ECONOMIC CHARACTERIS- TICS/COST HUMAN CHARACTERISTICS BIOLOGICAL/ECOLOGICAL CHARACTERISTICS COMMUNITY CHARACTERISTICS LEGAL CHARACTERISTICS TRAFFIC CHARACTERISTICS PHYSICAL CHARACTERISTICS CHEMICAL CHARACTERISTICS ELECTRICAL/ELECTRONIC CHARACTERISTICS MILITARY CHARACTERISTICS	A. NONTECHNICAL LEVEL B. ADMINISTRATION/ MANAGEMENT LEVEL C. TECHNICAL OPERATIONS D. ENGINEERING PRACTICE LEVEL E. APPLIED RESEARCH/ DEVELOPMENT/TESTING/ EVALUATION LEVEL F. BASIC RESEARCH LEVEL Y. OTHER SPECIFIC INFORMATION LEVELS 7. INFORMATION TYPE (Tipos de información) A. ACTIVITY UNIT NOTICE B. ACTIVITY UNIT REPORT C. ACTIVITY MANUAL D. ACTIVITY DATA SET E. PUBLICATION/ARTICLE F. BIBLIOGRAPHY	A. FEDERAL TRANSPORT AGENCIES B. OTHER FEDERAL AGENCIES C. REGIONAL OR STATE OR LOCAL TRANSPORT AGENCIES D. OTHER REGIONAL OR STATE OR LOCAL AGENCIES E. ACADEMIC OR RESEARCH INSTITUTIONS F. SOCIETIES OR ASSOCIATIONS G. BUSINESS OR INDUSTRY H. OTHER PUBLISHERS I. OTHER INFORMATION CENTERS J. INTERNATIONAL TRANSPORT AGENCIES/ CONFERENCES K. OTHER INTERNATIONAL AGENCIES/CONFERENCES	A. USA B. CANADA C. MEXICO/CENTRAL AMERICA D. SOUTH AMERICA E. UNITED KINGDOM F. FRANCE G. WEST GERMANY H. OTHER WEST EUROPE I. RUSSIA J. OTHER EAST EUROPE K. MIDDLE EAST L. SOUTHEAST ASIA M. OTHER ASIA N. JAPAN O. OTHER PACIFIC P. AFRICA Q. AUSTRALIA	MZ. US DOT MA. OST/ORT MB. OST/OPIA MC. FAA MD. FHWA ME. FHWA/BPR MF. FHWA/NHSB MG. FRA MH. UMTA MI. USCG NZ. OTHER US GOVT NA. MARAD NB. CFSTI NC. NBS ND. DDC NE. NASA OZ. US SOC/ASSNS OA. IRF PZ. US INFORMATION CENTERS PA. SIE PB. EI PC. HRB/NRC PD. MIC/NRC QZ. CANADIAN AGENCIES QA. DOT QB. TRANSPORT COMMISSION QC. CGRA RZ. UK AGENCIES RA. RRI/IRRD RB. TRAG RC. BRB RD. MIN TECH SZ. FRENCH AGENCIES SA. IRT TZ. INTERNATIONAL AGENCIES TA. IHRD TB. ECMT
OTHER SPECIFIC CHARACTERISTICS CHARACTERISTICS NOT SPECIFIED	Y. OTHER SPECIFIC INFORMATION TYPES	Y. OTHER SPECIFIC TYPES OF GENERATOR AGENCIES	Y. OTHER SPECIFIC GENERATOR LOCATIONS	YY. OTHER SPECIFIC INFORMATION SOURCES



SUBJECT CLASSES (Clases de Temas)

1. GEOGRAPHIC RANGES (Alcances geográficos)

A. URBAN TRANSPORT SYSTEMS
B. INTERURBAN TRANSPORT SYSTEM
C. INTERNATIONAL TRANSPORT SYSTEM
Y. OTHER SPECIFIC RANGES
Z. RANGE NOT SPECIFIED

2. TRANSPORT MODES (Modos de transporte)

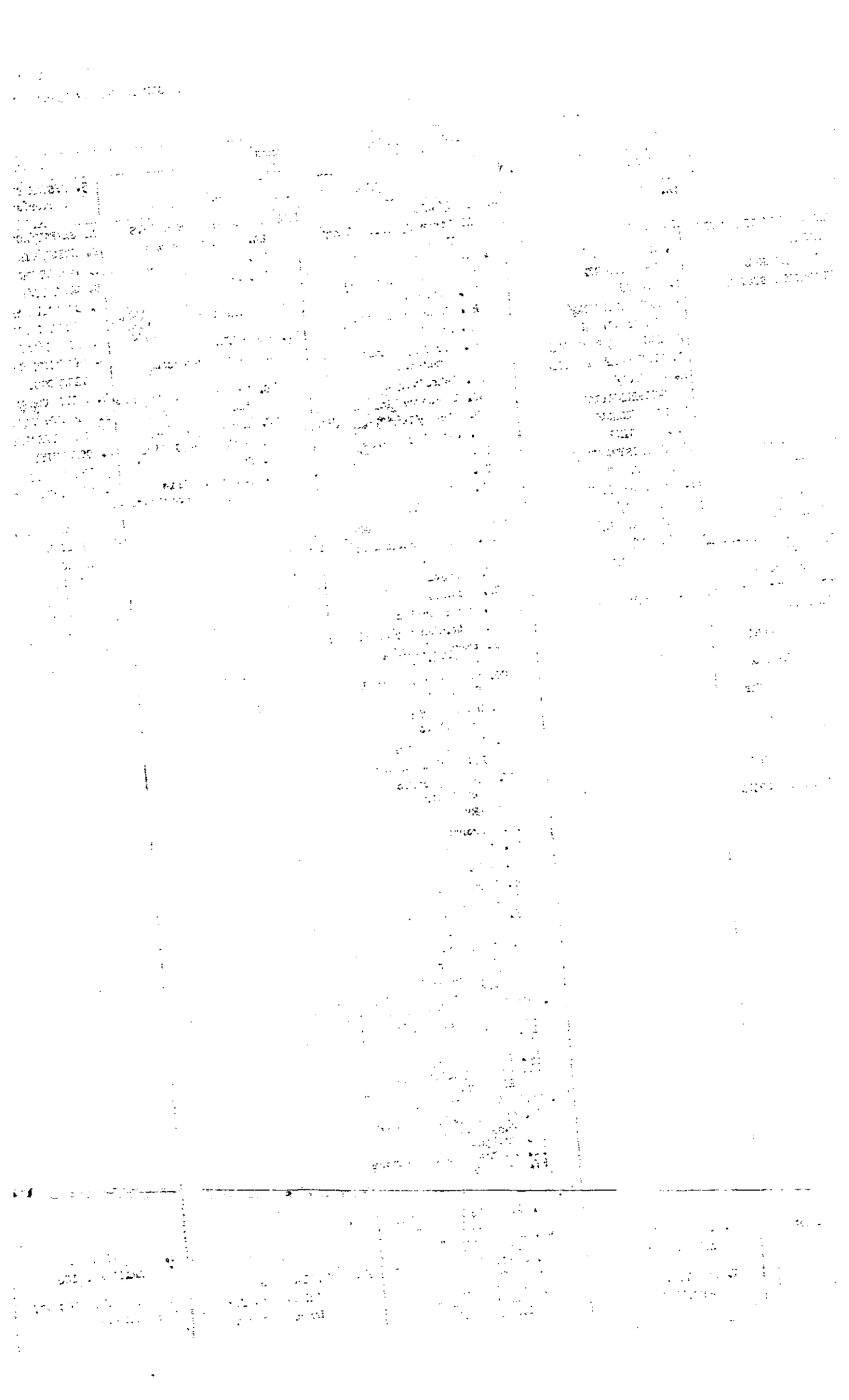
A. AIR TRANSPORT
B. HIGHWAY TRANSPORT
C. RAILWAY TRANSPORT
D. PIPELINE TRANSPORT
E. INLAND WATERWAY TRANSPORT
F. MARITIME TRANSPORT
G. MULTIMODAL TRANSPORT
Y. OTHER SPECIFIC MODES
Z. MODE NOT SPECIFIED

3. SYSTEM ACTIVITIES (Actividades de los sistemas)

A. ACQUISITION
B. PLANNING
C. SCHEDULING/PROGRAMMING
D. DESIGN
E. MANUFACTURING/CONSTRUCTION
F. HANDLING/PACKAGING
G. TRANSPORT OPERATIONS
H. CONTROL/COMMUNICATIONS
I. MAINTENANCE
J. FINANCING
K. ADMINISTRATION/MANAGEMENT
L. DEMONSTRATION
M. RESEARCH/DEVELOPMENT/TESTING/EVALUATION
N. EDUCATION/TRAINING
Y. OTHER SPECIFIC ACTIVITIES
Z. ACTIVITY NOT SPECIFIED

4. SYSTEM ELEMENTS (Elementos de los sistemas)

AZ LAND/WAYS
AY Other Specific Land/Ways
BZ. MATERIALS
BA. Soils/Aggregates
BB. Bituments
BC. Cement/Concrets
BD. Metals
BE. Marking/Coating Materials
BF. Adhesives/Seals
BG. Ceramics/Glasses
BH. Fibers/Textiles
BI. Plastics
BJ. Rubbers
BK. Wood/Paper
BY. Other Specific Materials
CZ. FACILITIES/STRUCTURES
CA. Runways/Pavements/Roadbeds
CB. Bridges
CC. Tunnels
CD. Pipes/Drains
CE. Interchanges/Terminals
CF. Storage/Parking Facilities
CG. Protective/Control Structures
CH. User Service Facilities
CI. Canals/Harbors
CJ. Industrial Plants
CY. Other Specific Facilities
DZ. VEHICLES
DA. Aircraft
DB. Automobiles
DC. Rail Vehicles
DD. Vessel/Ships
DE. Air Cushion Vehicles
DF. Special Ground Vehicles
DG. Buses/Trucks
DH. Containers
DY. Other Specific Vehicles
EZ. POWER/ENERGY CONVERSION
EA. Fuel/Power supplies
EB. Engines
EC. Motors
ED. Power trains
EY. Other Specific Power Elements
FZ. POPULATION AGENCIES
FA. Operators/Controllers
FB. Passenger/Users
FC. Freight
FD. Pedestrians/Spectators
FY. Other Specific Population/Agencies
GZ. ENVIRONMENTS
GA. Climatic Environment
GB. Physical Environment
GC. Socioeconomic environment
GD. Political Environment
GE. Psychological Environment
GY. Other Specific Environment
HZ INFORMATION ELEMENTS
HA. Signs/Markings
HB. Data/Documents
HC. Models
HY. Other Specific Information Elemen
IZ. EQUIPMENT/PROCEDURES
IA. Data Processing Utilities
IB. Maintenance/Constr tion Equipment
IC. Instrumentation
ID. Navigational Aids
IE. Sensors
IY. Other Specific Equipment/Procedur
JZ. HUMAN FACTORS
JY. Other Specific Human Factors



Topic 2

ROLES AND LINKS

Roles (Funciones) / Links (Nexos)	Role 8	Role 1	Role 2	Role 3	Role 4	Role 5	Role 6	Role 7	Role 9	Role 10	Role - or 0
	Primary Topics Principal Subjects (Temas básicos)	Inputs (Insumos)	Outputs (Salidas)	Undesirables Unnecessaries (Inconvenientes)	Present, Possible, and Later Uses (Usos)	Environments Solvents Media (Ambiente)	Independent Variables Causes (Variables independientes)	Dependent Variables Effects (Variables dependientes)	Passive Recipients (Receptores) Location (Ubicación)	Means of Accomplishment (Medios de acción)	Modifiers (Modificadores) Adjectives (Adjetivos) Names (Nombres) Bibliographic Data (Datos bibliográficos)
Terms (Términos)	Selection	Tax rate					Changes Rate Time Increasing	Selection Decisions	Structure Investments Research Development Completion Projects Programs		Corporate Alternative Short-term Long-term Applied Basic
Stylized Statements (Sinopsis)	A primary subject of discussion is the effect of changes in the corporate tax rate on the selection of alternatives among short-term and long-term investments, such as research and development projects. The time for completion of basic research as compared to the time required to complete applied research or development programs thus is a major factor in corporate investment decisions because of anticipated increases in corporate tax rate.										
Terms (Términos)	Evaluation Analysis				Maximizing Evaluation				Investment Return Costs Discounting Money Commitments Decisions	Pay-off Utility-function Prediction Discounting Rate Time-value Time	Alternative Varying
Stylized Statements (Sinopsis)	A primary topic of consideration is the evaluation of investment alternatives for the purpose of maximizing return to the investor, using payoff or utility-function and prediction of costs as well as varying rate of discounting future monies (the time-value of money) and the length of time of investment commitment. These analyses are to be used for evaluating alternative investment decisions.										
Terms (Términos) A			Parts				Rate Time-available	Performance Productivity	Production Feeding Operators Production lines	Conveyors Machines	Paced
B	Comparison Performance Productivity								Performance Productivity Operators Conveyors	Parts	Rigidly fixed Non-fixed
C	Design								Production lines	Time-available	
Stylized Statements (Sinopsis)	The primary topic of consideration is how feed rate and time available affect actual operator performance and productivity in production of parts on a production line paced by a conveyor or machine. The design of production lines based on time available is discussed. There is a comparison of operator performance and productivity using parts rigidly fixed to the conveyor and parts not rigidly fixed.										

Source: Training manual and workbook for use in abstracting and coordinate indexing - Training course prepared by John C. Costello, Jr. Battelle Memorial Institute, Rev. July 1964, Information System Research Div.

classification cannot be separated from the problem of retrieval of information, and the formulation of search questions and the realization of generic searches are difficult with less structured systems. Compared with more structured or more generalized schemes, faceted schemes can provide for more flexible combinations of terms.

33. The major disadvantage of faceted systems is that they are arduous to construct and "debug". Compared with less structured systems they cannot as readily incorporate new indexing terms. "In general, being intermediate in degree of structure, faceted schemes introduce more constraints than verbal indexes but have fewer search facilities than more structured codes." 15/

5. Relations of social and economic concepts for Latin America

34. In a classification scheme for social and economic documentation, there is a need for a system which is not so abstract and general as the MARC facet system, nor so specific as that of the Highway Research Board. What would be useful is an analysis of the basic elements of any social and economic system. These elements would be facets common to all the social and economic disciplines and would be common terms of reference facilitating the collection and comparison of materials about similar things from different fields. The final and most important test

---

15/ Vickery, op. cit., page 18.

/of the

of the worthwhileness of creating a facet is whether or not it facilitates the retrieval of materials relevant to policy areas in Latin America. 16/

35. Facets facilitate the retrieval of materials based on particular types of resources and problems and not just those that happen to have been recognized by specialists in the same discipline or subdiscipline as that of the user. Even within most fields there tends to be a great overlapping of themes which are treated by the major subfields of that discipline.

36. Thus the field of Public Administration is rather arbitrarily divided (in textbooks, university course structures, etc.), with no single characteristic used for defining the differences among the subfields. It is very likely that materials requested on "development administration" will be both overly inclusive and overly exclusive for application to concrete problems. While subdiscipline titles are often useful for categorizing materials on social or economic theory and methodology, it is increasingly evident in the social sciences that these titles have served as unfortunate straightjackets. There is, for example, much more in common between organization theory, political theory, and cost-benefit

---

16/ What is suggested here recognizes that "because of their essentially syntatic function (faceted), systems are not in fact substitutes, but rather complements, for the usual classifications and the most frequent case is that of a hierarchical organization superimposed on the facet". (1) Further it is acknowledged that "The faceted classifications seem to be somewhat less circumscribed in making changes than the hierarchical types, but this is partly an illusion. Ideas are the ultimate reality of man's existence because they determine how he will look at everything in his universe and what he will see or not see when he looks". (2) Thus, "In a facet classification, the validity of functional classes is related to the context or to the special subject which is treated, and, thus, they have their only value as classes in the given specific field... or restricted context". (3) It is precisely the context of the Latin American social or economic policy decision-maker that we want to capture. The above quotations are taken from Classification Research, ed. by Pauline Atherton. (Copenhagen, Munksjaard, 1965); (1) by J.C. Gardin, page 163, (2) by Phyllis A. Richmond, page 42, and (3) by J.C. Gardin, page 169.

/analysis than

analysis than the experts of those respective specialities have generally realized.

37. Let us consider some examples of facets which seem useful for separating materials according to policy considerations. As a start, time and space are dimensions common to all social and economic information. It would be very useful to have common terms referring to geographical space (e.g. names of nations, cities, etc.), economic space (e.g. urban, rural, interurban, subregional, etc.), and periods of time that the action takes place in the system. The concept of time can also be referred to in a different manner, according to type of time (e.g. sequential, cyclical, linear, etc.).

38. Types of development or change can be distinguished according to nature of the intervention of outsiders (e.g. autonomous, headstart, guided, controlled). Up to this point, these examples have been such that they seem to present little problem in their design. Once those index terms have boundaries which are carefully defined, there should be little trouble in applying them to any given situation.

39. When we consider the actions in a system which is studied, the actors, and that which is acted upon, the problem of setting up categories becomes much more complex. It is much more difficult to set up a set of index terms along these dimensions such that clear guides for classification of documents are apparent. For example, in describing social or economic action, it would seem useful to specify a sequence of activities commonly carried out in any social or economic project. We might want the documentalist to indicate if the document refers to planning, programming, fabrication, construction, operation, marketing, or final consumptive activity. However, is the use of limestone in a cement plant a fabrication or final consumptive activity? Is the use of cement in building a house a construction or a final consumptive activity? Perhaps sharp definitions can remove some of this ambiguity, and that which remains will be considered of little importance for our purposes. The Highway Research Board was able to come up with a useful set of index terms for "system activities" which satisfied their purposes and, unlike the HRB, our statements about facets are not intended to preclude the use of the more flexible synopsis

/form. We

form. 17/ We could create an index term "do not care" (meaning in this case "difficult to classify") and expect that the choice of descriptors in the synopsis will enlighten the user further. In this case, it would appear necessary that all documents identified as "do not care" be included in the output as far as that one facet was concerned. 18/ Of course many of these documents would not be included in the output, since they would be eliminated when tested in terms of other facets and descriptors.

40. As to what is acted on, a "policy-orientation" would seem to call for a separate reference to certain major products or groups of products in Latin America. Especially in the case of certain key imported and exported products, documentation on the production, mining, transport, and marketing of products will be of interest to key policy-makers. In the case of transport, variations of costs by product are influenced greatly by the type of packaging involved. This, however, may not be a convenient norm for grouping the products as they are described in the non-transport literature. Certainly the index terms here would have to be subject to constant development and it will be especially difficult to devise a list of products that will satisfy all persons. Some will prefer to have index terms based on origin or production (such as the LAFTA list), others index terms based on consumption. A completely unstructured list will not do, since this would be contrary to the purpose of providing a means by which administrators over certain "types of goods" can share their experiences.

41. At a very general level, the Classification Research Group (which worked up schemes on relations of concepts) has prepared schemes for things being acted on, categorizing objects according to levels of complexity (e.g. molecules - molecular assemblages - cells - organisms - human beings, etc.). 19/ This approach might be considered, applying it only to products.

17/ Another scheme is suggested in D. Langrieze's Classification of enterprise activities (London, British Institute of Management, 1956).

18/ This is easily arranged in the computer programming. However, if it is deemed desirable to single out documents dealing (e.g. a textbook) with all of the themes which are referred to by index terms of the facet, it is necessary to create a separate index term: "general". See Horst Koerner, "Short and flexible generic codes for mechanized information retrieval", in Classification Research, ed. by Pauline Atherton, op.cit., pages 275-6.

19/ Phyllis A. Richmond, "Contribution toward a new generalized theory of classification", ibid., pages 40-1

 An alternative

42. An alternative was devised by B.C. Vickery, in which products are differentiated according to their successive stages of manufacture (of pharmaceuticals):

"A Products
B Starting materials
C Substances to be extracted
D Reactions
E Physico-chemical operations
F Agents
G Properties
H Scale of operation" 20/

However, as J.C. Gardin notes:

> "It is clear that this order is exclusively funcitonal - the same substance or product may appear in facets: A, C, E, G, etc. - and is in consequence valid only for the field of study where these functions are themselves defined." 21/

43. Finally, the users may consider it useful to specify the major discipline involved, thus whether each document is primarily concerned with economic, financial, administrative, legal or infraestructure aspects of the subject being treated. They may even want a common breakdown of one or more of these aspects, but more likely this would be left for descriptors. The Highway Research Board has something similar under the heading "System Characteristics" - (Table 1).

44. Additional examples of facets for development problems and resources are the following (each with an example of several possible index terms):

Type of organization - (cooperative, private, public autonomous, public ministerial, etc.)

20/ Vickery, op.cit., page 56.

21/ J.C. Gardin, "Free classifications and faceted classifications", in Classification Research, ed. by Pauline Atherton, op.cit., page 165.

/Symptoms of

Symptoms of problem - (violence, poverty, political instability, inflation, technological gap, internal dependence, etc.)

Agents of change - (outside professionals, local professionals, military, bureaucrats, politicians, etc.)

Agents of reaction - (same as above)

Solutions proposed - (abandon, destroy, reorganize, leave alone, devaluate, etc.)

45. It would seem especially useful if the classification system designers, together with representatives of users, were to review the work of James G. Miller on taxonomies of biological and behavioral system components, or the work of others in this field, to consider relations of concepts which might be used as facets in the classification scheme. 23/ This literature can also be suggestive of definitions of the index terms included in each dimension (e.g. distinguishing "autonomous" from "headstart", "guided", and "controlled" development). A basic easy-to-read text is B.C. Vickery's Faceted Classification Schemes.

46. While the above literature may be suggestive of terminology and organizations of concepts, the usefulness of these concepts must be analyzed in terms of their ability to delimit the problems and resources of Latin America. For many Latin Americans, there is a substantial discrepancy between their own descriptions of their experience and the definitions in the models or paradigms used by the social scientists, professionals and government officials in their nations. The classification system designer has in his hands the power to help reorient the work of administrators of development and policy-makers in new directions, liberating them from the reliance on a limited number of variables and concepts by delivering references on social needs and effective techniques for meeting these needs in the Latin American reality. It is impossible to construct a useful system of classification without making commitments in social and economic theory.

23/ James Miller, "Living Systems", Behavioral Science 10 (October, 1965), pages 337-411. See also Talcott Parsons and Neil Smelser, Economy and Society: A Study of the Integration of Economic and Social Theory (London: Rovtledge and Kegan Paul, 1957), and Battelle Memorial Institute, System of Roles (Columbus, Ohio, n.d.).

/47. It is

47. It is presumed here that there has accumulated in Latin America a tremendous amount of experience with development strategies and programs. In many fields, there has been a great deal of experimentation with a wide variety of programs. The discovery and circulation of information on this experience with reforms, revolutions, and reactions is relatively new. The formation of new conceptualizations of this experience has become the preoccupation of a multitude of institutions in Chile, Mexico, Venezuela and other nations. The new interchange of ideas is so important, and the complaints are so intensive against the older methods of defining and conceptualizing Latin American needs, problems and resources, it would be a step backward to classify materials on Latin America according to classifications made by persons who are not familiar with these needs, problems and resources.

48. An example of a potential beneficiary of a regional or national documentation center in Chile would be an administrator of copper. How would the available materials be organized so that this administrator would have ready access to those most relevant to his needs? Naturally, the field of natural resources should have a descriptor "copper" for references to documents over the administration, processing, use and marketing of copper. However, this literature incorporates certain assumptions about copper which the Chilean administrator may want to avoid. As an administrator of a state enterprise in a society which wants to break with old patterns of dependence, he seeks new standards. He wants information relevant to his situation, whether or not it will be found among the specialized materials of natural resources. Would it occur to documentalists in other fields, such as transport, to classify their materials according to products, or types of products? Also, the Chilean administrator of copper desires that his own experiences be recorded and classified in such a way that other Latin American administrators with similar problems, whether or not they deal with minerals, can take advantage of them.

49. Once the facets which are preferred are chosen, the treatment of these facets in presently used documentation schemes should be examined and considered. It would be most convenient if the structure of index terms for one facet or another (e.g. products) adopted by the United Nations or one of its agencies was found to be appropriate for or compatible

with the needs of Latin American users. Compatibility with the index term structures of all the United Nations agencies would be impossible (since there is a lack of compatibility among them) [23], but compatibility with some of them certainly is desirable. However, convincing arguments have not been given in favour of the automatic imposition of one or another of these schemes and while the United Nations' goal of an integrated system should be appreciated, it should be recognized that this standard cannot and should not be applied rigidly. Also, the great amount of information per document (up to 10 000 bytes per document in four languages) that the United Nations Computer-assisted Indexing Project plans to record would be way beyond the needs or capacity of a Latin American documentation effort and this would have to be reduced. [24]

50. One condition to all of the above should be repeated. There should only be the amount of structure necessary to provide the user with the material relevant to the policy decisions which will be taken. That is, the facets which would be considered are those variables which are most important in policy considerations, where material will be drawn from several disciplines. If a variable is considered sufficiently important by this criterion, all materials of all disciplines should be indexed in terms of it, departing from a common set of index terms. The final scheme should be determined by consultation with professionals of all the relevant disciplines, as well as with governmental officials or their representatives.

6. Hierarchy and retrieval

51. There would have to be some hierarchization of the index terms included for each facet. The user requesting materials on "urban problems of cities in Venezuela" may not want to have to make a separate request for every possible detail, and yet may want material which deals in depth with the key problems, including "slum conditions in Caracas",

---

23/ G.S. Martini, "The United Nations System of Storage, Processing and Retrieval of Qualitative Information", monograph, (April, 1969), page 21.

24/ The United Nations system is being designed anticipating that other centers of documentation will do this. Martini, op.cit., page 16.

"transport problems in Maracaibo", and "public services in Ciudad Bolívar". In a hierarchization "slum" could be an economic space subcategory of "urban", and "Caracas" a geographical space subcategory of "Venezuela".

52. For mechanical processing, it is likely that shorthand notations (codes) would be used for the different index terms, whereby symbols are used instead of the entire terms. Additional codes could be used during retrieval to specify if the information should include first order (e.g. nations), second order (e.g. provinces), third order (e.g. cities), etc. terms in the hierarchy of terms. That is, a single code may be used to call not only for all the information on a certain subject with regard to Venezuela, but also the information on that subject with regard to all of the provinces and the major cities of Venezuela.

53. In the case of time, the problem is a bit different. The words "until", "after" or "between" could be used in order to relate the period of time referred to by the information in the document and the period of time of interest to the inquierer. "Until 1967" will be consistent with a request for "until 1969", but not for "between 1967 and 1979". Future dates would be used in the case of projections and prognostications.

7. The descriptors

54. Beyond the hierarchization deemed necessary, there should be a great deal of freedom among disciplines in selecting, adding and subtracting descriptors which would be included in the thesaurus. Our criteria for selecting facets may seem unduly restrictive, since it does not consider the classification of materials according to the needs of those carrying out basic research (theory, theses, etc.). However, it is these users who need a system with the least amount of structure possible, anticipating that they will experiment with a wide variety of definitions of problems and resources.

55. In choosing descriptors, it is useful to take maximum advantage of facets. Facets can be used in pairs or groups in order to designate a complex idea, avoiding the necessity of having as many descriptors.

/Instead of

Instead of a descriptor CACM (Central American Common Market), the fact that the economic space(Common Market) and geographical space (Central America) are referred to as facets insures reference to the CACM.

56. As in the present OECD system, words other than index terms or descriptors would be included in the synopsis, but searches could not be made on the basis of those additional words.

57. It would be useful if there were negotiations among the different disciplines on specialties to obtain some uniformity of terms, where the variations among terms would otherwise be slight or where there might be duplications. In fact, where it is decided that a certain type of content of all documents should be indicated, even though this content does not meet the specifications set out in the earlier discussion of facets, it would be treated the same as required facets. An example could be the names of international organizations.

58. To facilitate retrieval even more, the descriptors could be listed in the thesaurus in small groups, generally mutually exclusive in meaning. Within these small groups of descriptors, especially among descriptors which vary only slightly, albeit importantly, in meaning, the differences in meaning should be sharply defined. As noted above, this will be done for the Aligned Descriptor List. 25/

59. More inclusive, very broad categories for these groups of descriptors might be convenient, and certainly would be useful. This would mean a limited amount of hierarchization of titles, (broad categories, small groups, individual descriptors) in the thesaurus.

25/ There has been work on associative techniques, in computers, for retrieving documents relevant to a request even though the documents may not have been indexed with the exact terms specified by the request. Degrees of association among terms are used. This work should be investigated to see if it might be a handy, even if imperfect, device for helping the user find related materials described in varying terms. This is important because social science materials vary greatly in the amount of precision of their use of terms. Harold Borko (referring to the work of H.E. Stiles and J. Spiegel), "Research in Computer Based Classification Systems", in Classification Research, ed. by Pauline Atherton, op.cit., page 224.

/60. Rather than

60. Rather than having a facet of "system characteristics", or "disciplines", (e.g. administrative, legal, economic, financial, etc.), it may be preferable to group the descriptors into such broad categories. However, these categories should not be imposed, but rather they should be derived working from the specific to the general.

8. Example

61. Since the system we are suggesting would only be developed through consultation and study with the users, it is almost presumptuous to give an example. However, if some arbitrary assumptions are made, an idea could be given of what the output might look like.

62. First, an example of output using the Aligned Descriptor List is given here. 26/

"29068 1967 12783—
MALDONADO VA
PARIS. UNIVERSITE. INSTITUT DES HAUTES ETUDES DE L'AMERIQUE LATINE
LE MARCHE COMMUN LATINO-AMERICAIN - ASPECTS DE L'INTEGRATION ECONOMIQUE EN AMERIQUE LATINE.
PARIS, 1967, XIV, 293 P. TABLES. (ITS - TRAVAUX ET MEMOIRES, NO. 19.)
/THESIS/ON PROBLEMS OF AND /OBSTACLE/S TO /ECONOMIC INTEGRATION/ IN /LATIN AMERICA/ - INCLUDES A SECTION ON /EEC/ EXPERIENCE, AND COVERS /ECONOMIC DEVELOPMENT/ IN /LAFTA/ COUNTRIES, /TRADE/, /INDUSTRIAL STRUCTURE/S, THE /BALANCE OF PAYMENT/S, /FOREIGN INVESTMENT/, /ECONOMIC POLICY, /INDUSTRIAL POLICY/, /TRADE AGREEMENT/S, /NATIONAL PLANNING/, /MONETARY POLICY/, /FISCAL POLICY/, /CREDIT/ FACILITIES, ETC. /BIBLIOGRAPHY/ PP. 287 TO 289, /IMF/ AND /UN/ MENTIONED, AND /STATISTIC/S.
FREN"

63. Suppose we have several required dimensions, separated by two slashes, 27/ in this order: type of document, type of material, 28/ geographical space, economic space, time, products, system characteristics, and international organizations mentioned. We shall retain essentially the same descriptors, although this is by no means to suggest that they are considered adequate. The result could be something like the following: 29/

26/ From Centro Latinoamericano de Documentación Económica y Social, "Los descriptores y su uso", monograph (3 febrero 1971), page 2.

27/ The separator symbol actually used would be such that the number of bits required for the notation of search terms in computer storage is minimal.

28/ See footnote 8.

29/ The numbers in parentheses refer to relative importance where there is more than one referent. Where all have the same weight (approximately), a (1) is placed before the first referent of the group. I admit unfamiliarity with the thesis, which makes my example even more hypothetical.

/"//THESIS//

"//THESIS//BIBLIOGRAPHY (2), STATISTICS (2), TEXT (1)//LATIN AMERICA (1), EUROPE (2)//COMMON MARKET GROUP//UNTIL 1967//PRIMARILY INDUSTRIAL// (1) FINANCIAL, ECONOMIC//LAFTA (1), EEC (2), ECLA (2), IMF (3), UN (3)//

PROBLEMS AND OBSTACLES TO ECONOMIC INTEGRATION OF LAFTA; WITH SECTION ON EEC EXPERIENCE, COVERING (1)/TRADE/, /STRUCTURE OF INDUSTRY/, /BALANCE OF PAYMENTS/, /FOREIGN INVESTMENT/, /ECONOMIC POLICY/, /INDUSTRIALIZATION POLICY/, /TRADE AGREEMENTS/, /NATIONAL PLANNING/, /MONETARY POLICY/, /FISCAL POLICY/, /CREDIT FACILITIES/, ETC."

9. Citation Order

64. The mixed facet-descriptors scheme tentatively proposed here facilities understanding of the meaning of the major concepts in the documents in a number of ways. Because the use of the facets is obligatory, all documents of a certain class, with respect to a few very basic variables (the facets), will be assigned a common designation (the index terms). Furthermore, the weights attached to the terms aid the user in assessing the relative importance of the treatment given the different topics, not always evident even when the flexible synopsis form is used. Finally, the facets indicate the relationship of the terms, up to a point.

65. A big problem is the lack of a citation order, so as to give a sequence in which the dimension items or descriptors may be called. Although the lack of a citation order is not inherent in the use of facets, it has not been recommended here. In the example given above, the document has information on the effect or stand of LAFTA with regard to financial matters, but evidently does not deal with the finances of LAFTA. The citation order which would be relevant to the user is Finances-Latin America-Common Market, not Common Market-Latin America-Finances. However, since there is no citation order, calling the three terms (whether they are dimension items or descriptors) "Latin America", "Common Market", and "Finances" in the system described here will lead to a retrieval of this document even if the user wants information on the finances of LAFTA. One answer might be that the term "finance" is too generic in this case. One possibility is the separation of the field "budgetary and accounting" as a discipline, independent of "finance". Another possibility

/is that

is that the user could also call for "fiscal policy" or other such available descriptor to specify what he wants, in addition to calling "finance". There could be a different descriptor, like "finances of the organization", which would refer to documents on the finances of LAFTA.

66. It should be apparent, that, as much as possible, descriptors should be designed such that type "x" of "y" will not be confused with type "y" of "x" during the search process. The Battelle system of "links and roles" is constructed for overcoming this problem, but it is difficult to determine a most useful order of more specific facets for a heterogeneous user population. J. Mills recommends the following formula for determining the combination order of facets: 30/

> Whole Thing—Kind—Part—Material—Properties—Processes—Operations —Agents.

67. One possibility that should be considered is a two-step mechanical search, to be described later (under the theme "Using the Computer"), in which the user is able to make specifications about syntax.

10. Interpretation of information requests

68. A great deal of thought needs to be given to the proper amount and kind of intervention which the documentalist-librarian would make in handling information requests. Some systems, such as that of the Scientific Information Exchange (SIE) described above, use persons competent in the subject matter. There are a number of disadvantages to this, and the very cost it would entail seems to preclude its consideration in most cases.

69. In the case of a documentation center linked with a large professional organization (such as ECLA-ILPES or the OAS), the professionals could be a basic resource for the documentation center, while also being users. A representative of each area of specialization, or of groups of specializations, could be appointed as "technological gate keeper", or "contact" who would work closely with the documentalist-librarian in dealing with especially difficult requests. He would be informed enough about the use of terminals

---

30/ J. Mills, referred to in W. Wilfrid Lancaster, op.cit., page 13.

/and about

and about the classification schemes to be able to help out in this and a number of other ways. He would serve as a liaison between the documentation system and professionals other than himself in the area(s) of specialization for which he is responsible. He would make sure (a) that the information generated by his group and new publications in his subject area are communicated for reference in the documentation system; (b) that the classification scheme (the descriptors) is updated according to new concepts and variables in his subject area; and (c) that the maximum benefits from the documentation center to his professional group are being realized. 31/ Of course, he would be consulted during attempts to create classification titles common to all disciplines.

31/ See Bryan Yates, "The Pilkington Technical Comminications System: A Formalization of the role of the "Technological Gate Keeper", Aslib Proceedings (October, 1970), pages 507-10. In this regard, the following conclusions of a study by Janice Ladendorf are relevant, even though they refer specifically to scientists (basic research) and technologists (policy research): "Scientists and engineers can be provided with the finest of information services, but they will not necessarily use these services, nor will they engage in the painful effort of reading and, even worse, thinking about the materials which such services can provide them...Information flow always breaks down into communication between individuals and the increase or modification of knowledge stored in their brains. All that information science can attempt to do is to facilitate this process insofar as this is possible, and this necessarily requires quite a precise knowledge of user habits and needs... The reasons why a scientist or a technologist prefers to consult a knowlegeable friend rather than the literature are fairly obvious. First, they share the same professional language and second, the information he receives has been refined, adapted, and evaluated to fit his specific problems. This type of communication therefore involves much less painful effort on the part of the inquirer. In technology, oral communication plays even a more crucial role than in science, since technologists as a group make much less use of literature than do scientist... Unfortunately, the formal information system - specifically libraries or information centers - is usually considered to be an information source which requires a high degree of effort on the part of the user." Following from these conclusions, Miss Ladendorf suggests the formalization of the "technological gate-keeper" role. Janice M. Ladendorf, "Information Flow in Science, Technology and Commerce: A Review of the Concepts of the Sixties", Special Libraries (May/June, 1970), pages 215-221.

/70. The relevance

70. The relevance of the information which is retrieved can be greatly increased by giving special attention to the formulation of request statements by users. There are several simple devices by which the documentalist-librarian can improve his understanding of the particular interests of the user. For example, it is useful to ask the user to indicate the titles of known relevant articles. Naturally, other articles classified with the same index terms as those used for the known articles would be retrieved. Also, the documentalist-librarian could present a display of the proposed formulation of his request with the index terms. The requester would have an opportunity to make suggestions at that point. This follows the dictum that "On the whole, it seems that a requester is able to define his need more precisely the later in the total retrieval operation that the interaction /between the user and the system/ takes place." 32/

11. Using the computer

71. Undoubtedly, the task of computer programming will be very complex. Beginning in the very early stages of classification system design, there is the need for continuous consultation with an expert on the options and requirements presented by usage of computers for information retrieval. Before any work is done on programming itself, this expert should investigate the existence of package programmes which would obviate the need for starting the programming from scratch, and would advise on the costs and advantages of using different types of storage and processing devices. The expert would suggest means by which the documentation center would prepare for future adaptations to more complex and advanced systems in order to efficiently handle greater volumes of interrogations, and of up-dating of the materials being stored. 33/ All of these considerations

---

32/ F. Wilfrid Lancaster, op.cit., page 190.

33/ Gunnar A. Berggren, inter-regional adviser on computer methods, "Report of Visit to Santiago, Chile, 8-30 April 1970", page 3.

/are interdependent

are interdependent with the design of the classification scheme. It is not a simple matter of a translation of the classification scheme into codes and the derivation of a programme to process the coded material. The selection of equipment and the design of the classification scheme, the coding system, the input format, and the output format need to be a singular operation. Even while sufficient funds for just terminal operations are lacking, 34/ the system should be designed for eventual computer operation.

72. There have been a number of attempts to develop package or "general" programmes for the different types of classifications schemes. Thus for the faceted classification scheme, a pure scheme (for facets only) is "formulated in terms of a very simple model which in some ways constitutes the archetype of all faceted classifications, independent of their field or their particular contents. From the notion of the model we pass easily to that of a general programme for the recording and automatic searching of data expressed in accordance with the norms of the model, a programme which is applicable to the exploitation of all faceted documentary systems, whatever the field may be". 35/ The computer systems expert would review the work which has been done on these package programmes and would report on the feasibility of adapting them for use in the Latin American documentation scheme.

73. It seems reasonable to expect that one such package programme can be obtained and merged with a package programme (making the necessary alterations, of course, of for a compatibility of languages) used for flexible synopsis systems. Most documentation systems, as J.C. Gardin has observed, include a mixture of classification schemes and automatic treatment of such systems involves the synthesis of programmes with different international organizations and exploratory methods. 36/

34/ As to CLADES, see the budget prepared by Mr. Robert Mantz, "Need for and viability of a United Nations Computation Center in Santiago", (21 June 1970), page 27. Also see G.K. Thompson, "Some Cost Estimates for Bibliographical Searching in a Large-scale Social Sciences Information System", Information Storage and Retrieval, Vol.6, Pergamon Press, 1970), pages 179-186.

35/ J.C. Gardin, op.cit., page 167. He reports on specific experiments underway in 1965 with a packaged programme for faceted schemes.

36/ Ibid.

/74. A major

74. A major advantage of faceted classification schemes is that information retrieval is facilitated. The large amount of theorizing and experimentation with generic codes (codes for handing hierarchical sets of classification titles) for computer retrieval can be drawn upon. In this regard, it is interesting that the proposals in this essay are consistent with a list of suggestions regarding classification scheme design by Horst Koerner, in an article on mechanized information retrieval. 37/

75. One of Koerner's suggestions is particularly interesting:

> "Use a two-step search, disregarding syntax during the first, coarse search. Look up the retrieval document numbers in a second file, which includes syntactical information (meaningful sequence, phrasing or grouping of terms, role indicators, interfixes, etc.) for comparison with the syntactical search description. Disregarding syntax allows very fast processing of the whole file. This ordering requires that any meaning attached to the sequence of terms has to be given up. In the second step, when only a few descriptions have to be compared, all the syntax desired can be used. For the second step a large direct (random) access file, which could be quite slow, would be ideal. In the second file could be stored, in addition to the description terms, not only syntactical information but also names, etc., and very specific, rarely used terms for which the classification tree of the coarse file cannot be further subdivided."

Rather than having a second file as Koerner suggestions, there should be a second programme to eliminate references on the basis of special details on syntax to be provided by the user (on standard forms, of course). That is, after a first pass, retrieving references to documents which are described with the same index terms as those posited by the requester, there would be a second pass to eliminate many of these references. In terms of computer operating costs and time, it is worthwhile to engage in additional processing of the information in the computer in order to reduce the amount of unwanted information which is printed out.

76. Key considerations in the design of a classification scheme are the amount of computer storage and the amount of scanning time which are required. The amount of computer space occupied for each document depends

---

37/ Koerner (with comments by Salton), op.cit., pages 268-289. The importance of some of Koerner's suggestions depends on the amount of pressure there is on computer memory space.

/inversely on

inversely on the amount of structure and directly on the number of index terms and descriptors. The more structure provided by the facets, and the more structure given to the list of descriptors, the less the number of irrelevant combinations of terms that will be sought. What are called "multiple look-up techniques" for mechanized information retrieval could be used, the searches being based on combinations of the facet terms. Such systems take advantage of the quicker search time opportunities provided by disks or other Direct Access Storage Devices (drums, data cells, magnetic card cartridges, etc.).

77. Some computers are less appropriate than others for rapid information retrieval. Some multiple-use third generation computers, designed for sequential scanning, have impressed users with the speed by which search is carried out for references in a flexible classification scheme. However, fractions of this search time can be obtained by using other computers, including those which have fixed head disks. This is especially the case when these machines are combined with a more structured classification scheme.

78. The development of a computer-based documentation system is a long-term and expensive project. A careful study of the experience of other documentation centers in this regard can save Latin American centers a certain amount of tears and cash, but it will undoubtedly be necessary to develop the computer operation on a phase by phase basis (see Annex I for an example). This is necessary simply because the number of interrogations, experience with the documentation effort, and most of all the available financial resources, will be insufficient to support a complete system right from the beginning.

/Annex I

Annex I

EVOLUTION OF THE BIBLIOGRAPHICAL RETRIEVAL SYSTEM
OF THE INTERNATIONAL LABOUR OFFICE

	Number of records available	Average search time	Average users' waiting time
Phase 0 (1964) Planning and experimentation	-	-	-
Phase 1 (January 1965-April 1969) Punched cards, including inverted descriptor file	0-28 000	up to 8 hrs.	2-3 days
Phase 2 (May 1969-) IBM 360/30 batch processing	28 001-	90 sec.	less than 24 hrs.
Phase 3 (August 1969-) IBM 1050 terminal on-line throughout the day	31 000-	instantaneous	few minutes to few hours (depending on print out)
Phase 4 (1970-) Visual terminals in various locations, 2314 disks	40 000-	instantaneous	few minutes to few hours (depending on print out)
Phase 5 (1971/72?-) Inter-agency network	100 000-	instantaneous	few minutes to few hours (depending on print-out)

Source: "Some Cost estimates for Bibliographical Searching in a Large-scale Social Sciences Information System", by G.K. Thompson. Reprinted from Information Storage and Retrieval, Vol. 6, pages 179-186. Pergamon Press 1970.

Note: Some 50 000 bibliographical records of 500 characters average length can be stored on one 2314 disk, together with the inverted descriptor file.